

FAIRLY PARTITIONING RESOURCES WHILE LIMITING THE MAXIMUM FAIR SHARE

ABSTRACT OF THE DISCLOSURE

Resource requests from a plurality of schedulable entities are scheduled while limiting
5 the maximum and minimum quality of service allocated to each schedulable entity. The resource
scheduler of the present invention requires less memory to maintain state information than
existing rate-controlling schedulers, and is thus more easily scalable to large numbers of users.
The resource scheduler also schedules resources fairly among competing schedulable entities. A
fair-share scheduling algorithm is used by a resource scheduler to select resource requests to
service. A rate controller checks to ensure that servicing the selected request will not cause the
associated user's maximum quality of service to be exceeded. If the maximum quality of service
will not be exceeded, the virtual time used in the scheduling algorithm is incremented, and the
request is serviced. If the maximum quality of service will be exceeded, the virtual time is still
incremented, but the request is not serviced and remains pending.

0020005252900
15